

CDSCO-IndiaAI Health Innovation Acceleration Hackathon

1. Introduction

IndiaAI, an Independent Business Division (IBD) under the Digital India Corporation (DIC) of the Ministry of Electronics and IT (MeitY), is the implementation agency of the IndiaAI Mission, which aims to democratise AI's benefits across all strata of society, bolster India's global leadership in AI, foster technological self-reliance, and ensure ethical and responsible use of AI.

As part of this Mission, the IndiaAI Application Development Initiative (IADI) aims to promote the development, deployment, and adoption of AI applications in critical sectors that have the potential to catalyse large-scale socio-economic transformation.

Towards this, IndiaAI is collaborating with Central Drugs Standard Control Organisation (CDSCO), the Central Drug Authority for discharging functions assigned to the Central Government under the Drugs and Cosmetics Act, 1940. It exercises regulatory control over the import of drugs, approval of new drugs and clinical trials, meetings of Drugs Consultative Committee and Drugs Technical Advisory Board, Serious Adverse Event (SAE) monitoring, and approval of certain licenses as Central License Approving Authority.

CDSCO reviews and evaluates an enormous volume of submissions, for new drugs, biological products, and clinical trials. These submissions demand intensive and minute review of critical data, including clinical trial data, manufacturing quality control and safety information. The current process remains manual and document-intensive, leading to systemic inefficiencies and delays. It significantly impacts the timely introduction of new drugs and medical devices constraining public health access and innovation across India's healthcare sector. Further, CDSCO also undertakes assessment of SAEs to ensure effective resolution as per CDSCO guidelines and standards.

With the aim of streamlining the regulatory review process to advance the future of India's healthcare, the IndiaAI-CDSCO Hackathon seeks to leverage Artificial Intelligence (AI) and Machine Learning (ML) to promote innovation by enabling faster, consistent, and efficient approvals.

The Hackathon aims to bring digital transformation to reduce administrative burdens for reviewers, improve the depth and quality of the entire review process, and expedite critical regulatory decisions, ensuring that life-saving innovations reach citizens faster. Further, to ensure strict compliance of AI/ML solutions with privacy and ethical guidelines, the Hackathon will focus on development of an AI-powered anonymisation tool capable of processing both structured (forms, tables, databases) and unstructured data (documents, PDFs,

transcripts). Such anonymisation ensures compliance with the DPDP Act 2023, National Digital Health Mission (NDHM) Health Data Management Policy, Indian Council of Medical Research (ICMR) Ethical Guidelines, and CDSCO standards, while enabling the responsible use of data for innovation in regulatory innovation.

2. Hackathon Process and Participation Details

I. Stage 1 (Application and Virtual Hackathon):

a. Participation Requirement:

- Participants may participate as individuals or in groups as per the eligibility criteria detailed in Section 7.

b. Participation Process:

- All participants are required to utilize the AIKosh portal to access the application form.
- A team leader will have to individually apply for the Hackathon on the AIKosh portal by clicking the submit link.
- After the initial sign-up, the Team Leader should list all the Team Members under the Management Team and complete all organization details.
- Team Leaders will create an account on GitHub to upload Source Code. Next, the Team Leader shall paste the link for their GitHub repository in the application form.
- Additionally, the Team Leader will have to answer all additional questions, including uploading documents, and click 'Submit'.
- The Hackathon under Stage 1 would take place virtually over 45 days from the date of launch.
- Any edits to the Source Code post the final date of submission will lead to immediate disqualification of the application.

II. Stage 2 (Shortlisting and On-premises development):

a. Participation Requirement:

- Up to top 10 teams will be selected for an on-premises development round.
- Up to 3 members per team will be provided travel and accommodation support to participate.
- *IndiaAI reserves the right to modify the number of qualifying solutions to ensure competition and operational efficiency.*

b. Participation Process:

- After shortlisting in round 1, the second round will be organised as a 5-day on-premises round at CDSCO office, New Delhi. Details of the location and logistics will be communicated via email to the shortlisted teams.
- To ensure the confidentiality and security of the datasets, teams will be granted access to the required data through a secure server to address the problem statement in Section 3 pursuant to signing a Non-Disclosure Agreement (NDA).

3. Problem Statement

I. Problem Statement: (Stage 1) - AI-Driven Regulatory Workflow Automation and Data Anonymisation

Expected Output: A comprehensive solution that integrates the following features:

- AI-powered data anonymisation tool (web or API-based):** Detect and classify Personally Identifiable Information (PII) or Protected Health Information (PHI) in both structured and unstructured data using a hybrid rule-based and NLP approach for contextualised anonymisation. Implement a two-step process: de-identification (Pseudonymisation), replace identifiers with secure tokens and irreversible anonymisation, generalise and normalise sensitive information, in compliance with Indian and International data protection laws and guidelines (DPDP Act 2023, NDHM, ICMR, and CDSCO guidelines) with the ability to streamline CDSCO's regulatory evaluation using AI to achieve faster, more consistent approvals and enhance public health access.
- Document Summarisation:** Extract and synthesise critical regulatory information from diverse, high-volume source documents and convert it into a precise, standardised text format for reviewer consumption. The system must handle three distinct types of source material: application data corresponding to each item across multiple checklists on the SUGAM portal, SAE case narration to facilitate efficient case-resolution workflows, and meeting summaries to synthesise key decisions, actionable items, and next steps from lengthy meeting transcripts or audio files. The final output across document types must be a concise, standardised text format enabling due diligence and decision-making by reviewers.
- Completeness Assessment:** Verify the completeness, consistency, and accuracy of administrative sections, mandatory forms, and regulatory checklists for each clinical approval application, as well as for SAE reports with detailed flagging of missing/inconsistent fields.

- iv. **Classification Tool:** Classify cases by severity (death, disability, hospitalisation, others), detect duplicates, and help officers with prioritisation for efficient review.
- v. **Document Comparison:** Identify and highlight substantive and specific changes between different versions of an applicant's filing. A visualisation or report detailing significant changes in text, data, or tables between document versions.
- vi. **Inspection Report Generation:** Convert unstructured and handwritten observations gathered during site inspections into standardised, formal reports conforming to CDSCO inspection templates.

Applicants are suggested to integrate all aforementioned features with a particular focus on the first four features in their submitted solution as part of the application process.

II. Problem Statement (Stage 2): AI/ML to streamline CDSCO's review process

Objective: Refine and test the solution developed in Stage 1 to further streamline CDSCO's regulatory approvals using AI to achieve faster approvals and enhance access to healthcare innovations.

Expected Output: The required solution must be a comprehensive end-to-end system, incorporating all features specified in Phase 1. The system will be refined and trained using CDSCO-provided datasets to measurably enhance the efficiency of the regulatory approval process and is required to be integrated with the SUGAM and MD Online portals.

A **final project report** summarising outcomes and recommendations.

4. Expected Deliverables from Participants

I. Model Code and Documentation (GitHub):

- Explanation of the key methodology and steps taken in model development.
- Steps to grant access to your GitHub repository:
 - Go to the main page of your GitHub repository.
 - Click on the 'Settings' tab in the menu bar.
 - In the left sidebar, select 'Collaborators'.
 - Under the 'Manage Access' section, click on 'Add people'.
 - In the text field, search for 'indiaaihackathon25' and add it as a collaborator.

- The team shall paste the link for their GitHub source code in the application form.

II. Project Report (PDF):

- Key Findings from Your Analysis
 - a. Detection Methodology: Detailed explanation of the approach (e.g., named entity recognition models, regular expressions used, extractive vs. abstractive NLP, transformer models, etc.). Detailed rules or machine learning logic used to verify completeness, consistency, and accuracy across forms and checklists. Description of the algorithm used to identify and highlight substantive and specific changes between document versions (e.g., semantic comparison vs. simple diff).
 - b. Anonymisation Report: Sample output showing raw data and the corresponding anonymised data, including the secure tokens used for pseudonymisation. Explicit documentation on how the two-step process (de-identification/Pseudonymisation and irreversible anonymisation/generalisation) is achieved.
 - c. Flagging Mechanism: Demonstration of the mechanism for detailed flagging of missing/inconsistent fields in both clinical applications and SAE reports. Methodology used to detect duplicate cases.
 - d. Classification Criteria: Clear definition of how cases are classified by severity (death, disability, hospitalisation, others).
 - e. Specific strategy for handling the three distinct source materials like Application Data (Checklists), SAE Case Narration and Meeting Transcripts/Audio Files.
 - f. Demonstration of the final concise, standardised text format for each source type. Explanation of the logic used to assist officers with prioritisation for faster review. Sample output of the visualisation or report detailing significant changes in text, data, and tables.
- Evaluation of the Model
 - a) Report how the solution performed against the metrics provided.
 - b) Note key limitations or areas that require further refinement.
- Implementation Plan

- a) Suggest improvements, additional data needs, and how the solution could be extended or scaled in the future.
- b) Briefly mention how data security and retrieval mechanisms could be incorporated in later phases.

5. Evaluation Process

The evaluation process of the Hackathon would be overseen by a distinguished panel of jury members comprising experts from the fields of machine learning, data science, and cyber security including senior officials from CDSCO. The jury would rigorously assess each submission based on predefined criteria to ensure a fair and comprehensive evaluation. The evaluation will ensure equitable weightage is given to both the Technical and General parameters.

Stage 1

- **Initial Screening:** Submissions would undergo an initial screening to ensure compliance with submission guidelines and basic functionality.
- **Technical Evaluation:** The jury will conduct a detailed technical evaluation of the applications received.

Stage 2

- Based on the initial evaluation, up to 10 teams will be shortlisted to refine, integrate required features and pilot their solutions at CDSCO premises.
- Based on the results of Stage 2, the winning teams of the Hackathon may be selected.
- Prizes would be awarded to the top three teams whose solution meets the jury's satisfaction. A special prize would be awarded to women only teams, if any.
- In addition to the prize money, the first prize winning team will secure a chance to receive a work contract of up to INR 50 lakhs for a period of one year to develop, integrate and deploy their solution on a chosen secure Cloud Environment.
- If no solution meets the required standards, consolation prizes may be awarded based on the jury's discretion.
- The jury's decision would be final and binding.

IndiaAI and the jury reserve the right to modify the number of qualifying solutions at any stage to ensure competition and operational efficiency.

6. Opportunity for Applicants

- **Opportunity to Build for the Nation:** Contribute to developing innovative solutions that address critical challenges faced by the country, making a direct impact on society.

- **National Recognition:** Gain visibility and recognition from government officials, industry leaders, and peers for your contributions and innovative ideas.
- **Networking Opportunities:** Connect with like-minded innovators, potential collaborators, and key stakeholders in the tech and innovation ecosystem.
- **Exposure to Real-world Challenges:** Work on pressing issues faced by the nation, providing practical experience and a deep understanding of real-world problems.
- **Support for Implementation:** The winning solution will get potential support in scaling and implementing the solution at a national level, bringing your ideas to life.
- **Prizes and Incentives:**
 - **First Prize:** Up to INR 10 Lacs
 - **Second Prize:** Up to INR 7 Lacs
 - **Third Prize:** Up to INR 3 Lacs
 - **All Women Team Prize, if no women team in top 3:** Up to INR 5 Lacs

7. Eligibility

- **Indian Company:** The team may be an Indian company registered under the Companies Act, 2013. An Indian company must have 51% or more shareholding by Indian citizens or persons of Indian origin.
- **Start-up:** The team may be a start-up as defined in the latest notification by the Department for Promotion of Industry and Internal Trade (DPIIT), accessible at Startup India.

The participating entities must demonstrate development and deployment experience of relevant solutions with previous engagements in the private and public sector.

8. Timeline

#	Activity	Timeline
1	Launch Date	TBC
2	Last Date for Online Submission	TBC
3	Announcement of Results of First Round	TBC
4	Round 2 (On-premises, New Delhi)	TBC

5	Announcement of Winning Entity and pilot	TBC
---	------------------------------------------	-----

9. Intellectual Property Rights

The models submitted or awarded will become the property of IndiaAI and CDSCO, including all intellectual property rights to their underlying methodologies and innovations, and the participants shall be deemed to have given their no objection/consent for the same and shall also remain bound by the terms of a Non-Disclosure Agreement (NDA) with respect to such work. The participants agree to provide a No Objection Certificate as an author in favour of IndiaAI and CDSCO for the purposes of IPR registration and ownership rights, as and when required by IndiaAI and CDSCO.

10. Terms and Conditions

- a. All participants and team have to be eligible (See Eligibility Criteria) to participate.
- b. The participants will ensure code is free from viruses, malware. The participants will not use this contest to do anything unlawful, misleading, malicious, or discriminatory.
- c. The solution must not violate/breach/copy any copyrighted or patented concepts in the AI market.
- d. For Stage 1, the applicants are required to utilise publicly available datasets. For Stage 2, anonymised or synthetic datasets provided by CDSCO will be confined to the chosen secure Cloud Environment.
- e. The solution must not violate any data protection and governance regulations and policies.
- f. The solution must be in adherence with related cybersecurity and data protection standards and guidelines of the Government of India.
- g. Solutions must adhere to ethical principles and guidelines for the development, deployment and use of AI technologies, including fairness, transparency, accountability, and non-discrimination.
- h. The developed solution/product will be deployed in the chosen Cloud Environment and used for Union/State/UT government entities.
- i. Any new enhancements, features or innovation should be released on the chosen Cloud Environment.
- j. The winning entity shall receive a work contract of a fixed amount to support the solution development and deployment for at least one year from the go-live period. The support includes manpower and operational requirements for end-to-end development, deployment, maintenance, and bug fixing across the entire application.
- k. The winning entity shall submit progress-cum-achievement reports at quarterly intervals on the progress made on all aspects of the project

including expenditure incurred on various approved items during the one-year contract period.

- l. The winning entity is not allowed to entrust the implementation of this project for which work contract is received to another institution and to divert the award received from IndiaAI as assistance to the latter institution.
- m. The winning entity should not enter into collaboration with a foreign party (individual/academic institution/industry) in execution of this project without prior approval of IndiaAI.
 - a. The winning entity is free to market the product to any entity outside the Union/State/UT Government Organisations of India.
 - b. In case of any dispute on any other matter related to the project during the course of its implementation, the decision of the CEO, IndiaAI shall be final and binding on the winning entity.
 - c. IndiaAI reserves the right to modify the terms and conditions of this challenge for operational feasibility and compliance to rules and regulations of the Government of India.

10. Plagiarism and Ethics

- a. Participants are expected to uphold the highest standards of ethics and integrity throughout the Challenge.
- b. All work submitted must be original and developed by the participant or their team.
- c. Plagiarism, or the use of someone else's work without proper attribution, is strictly prohibited and would result in immediate disqualification.
- d. Participants must ensure that their solutions are proprietary and not copied from existing projects or code repositories.
- e. Moreover, the use of any external resources or pre-trained models should be clearly cited, and proper permissions should be obtained where necessary. Adherence to these ethical guidelines ensures a fair and competitive environment for all participants.
- f. By registering for this Hackathon, participants are giving an undertaking to adhere to all plagiarism and ethical guidelines set forth by the IndiaAI.

Annexure I: Evaluation Parameters

I. General

	Parameter	Description
1	Approach Towards Problem Solving	Product Idea, Degree of Innovation, Simplicity of Final Solution, Uniqueness of Idea, Novelty of Approach.
2	Solution Technical Feasibility	Product features, Scalability, Interoperability, enhancement & expansion, Underlying technology components & stack and futuristic orientation, and System
3	Product Cost	Initial cost and manpower cost after stage 2, other Potential costs to build, deploy and maintain the Product for one year from go-live.
4	Team Ability & Culture	Prior Experience of the entity in developing and deploying similar solutions in private and public domains specifically healthcare, Team Leader's Effectiveness (i.e. Understanding of subject matter, Ability to guide, Ability to present idea), Ability to scale up and market the product, Growth Potential of Organisation.
5	Adherence to Responsible AI Principles	Safety and Reliability, Equality, Inclusivity and Non-discrimination, Privacy and Security, Transparency, Accountability, Protection and Reinforcement of positive
6	Adherence to Data Policies and Cyber Security Guidelines	Adherence to applicable Government of India policies, guidelines, regulations on Data Governance and Cyber Security.

II. Technical

	Parameter	Description
--	-----------	-------------

1.	D a t a preparation	Participant has <ul style="list-style-type: none"> - explored the data and removed unnecessary columns. - checked if there is any skewness in the data and tried to mitigate it. - performed a stratified train-test split successfully to create train & test datasets.
2.	Model Building	<ul style="list-style-type: none"> - Participants have performed the required cross-validation and have built different models on raw data. - After evaluation on the raw dataset, Model hyperparameters are tuned using correct principles and the approach is explained clearly. - A reasonable number and variety of different models are attempted, and the best one is
3.	M o d e l Evaluation	Model evaluation is conducted using an appropriate metric.
4.	C o d e readability and conciseness	Efficient, concise code is written. The code is well documented, and the analysis is explained in the report format with findings from the dataset.

5.	<p>T e c h n i c a l R o b u s t n e s s</p>	<ul style="list-style-type: none"> • Text Recognition tasks: Character Error Rate (CER) on standard public benchmarks for scanned documents, such as the ICDAR 2019 ArT dataset or the SROIE dataset, or equivalent. • Anonymisation: specify the solution’s k-anonymity, l-diversity, t-closeness based on public benchmarks. • Key Information Extraction: the strict, entity-level F1-score on the FUNSD (Form Understanding in Noisy Scanned Documents) benchmark, or equivalent. • Summarisation task: ROUGE-1, ROUGE-2, and ROUGE-L scores on a standard summarisation benchmark like CNN/DailyMail/Xsum or equivalent. Also, BERT Score. • Segmentation tasks: mean Intersection over Union (mIoU) for segmenting document regions (e.g., tables, text, figures) on the PubLayNet or DocLayNet benchmarks or equivalent. • Classification, and comparison tasks: Macro-F1 and Matthews Correlation Coefficient (MCC). Provide confusion matrix, a table providing a detailed breakdown of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). • Latency: Time taken to process and anonymise a standard document/dataset. • Additional Criteria: Any other metrics as agreed upon by jury members.
----	---------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Annexure II: Application Form:

Post Login (Application Form)

1. Application Form (Main Section)

Section 1: Organisation Information

- **Applicant Type:** Drop Down - a) Startup; b) Company (Mandatory)

Team Leader Information

- Full Name*
- Designation*
- Core Expertise Areas*
- Email Address*
- Phone Number*
- LinkedIn Profile*

Organisation Information

- Name of Organisation *
- Registration Number*
- Date of Incorporation *
- Number of Employees *
- Core Function of Organisation*
- Address*
- Website*
- Prior Experience in relevant project implementation and research work (in 200 words):* Describe relevant AI solutions designed and developed, technologies used, and outcomes, relevant publications or patents, etc.
- Experience Collaborating with Government Entities (in 200 words):* Specify partners, nature of engagement, and key results achieved

Team Member Information

- Name
- Role
- Email
- LinkedIn Profile

Section 2: Project Proposal Architecture details

- Solution Description (in 200 words)

Provide a comprehensive overview of your AI solution, including:

- o Functionality*
- o Features*
- o Core AI technologies used*
- o Training and validation data used, highlighting data provenance, coverage, size and quality*
- o Process and strategies adopted for data preparation, AI/ML technologies selection, training, hyper-parameter tuning, refinement, solution monitoring*

and enhancement

o Solution replicability across multiple sectors for relevant use cases"

- Source Code URL:* *Link to GitHub repository with information on source code used to build, train, and test submitted AI solution, report, and readme file in a private repository.*
- Solution Report (in PDF format)*: *Report with details outlined in Section 4 of the submission guidelines.*
- Model Card & Architecture: Upload a model card covering security and privacy aspects, and providing the solution architecture diagram.

Section 3: Supporting Documents

- Certificate of Incorporation/Legal Registration (PDF):* Attachment - one word/PDF file, max 10 MB
- Certificate of Recognition (for Startups) (PDF): Attachment - one word/PDF file, max 10 MB
- Solution demo video (2-3 minutes): one MP4/MPEG/MOV file, max 200 MB

Section 4: Declaration

- I/We declare that all the information provided in this application is true and complete to the best of our knowledge. I hereby also declare that I am authorised by my company/startup/organisation to participate in this Challenge, sign all the documents and agreements related to the Challenge and to commit resources.
- I/We confirm that we will abide by the conditions mentioned in the Challenge document in full and without any deviation.
- I/We shall observe confidentiality of all the information passed on to us in the course of the Challenge and shall not use the information for any other purpose than the current Challenge.
- I/We confirm that we have not been blacklisted/banned in the last three years by any State/Central Government organisations/Firms / Institutions/ Central PSU / PSE.